DOCUMENT RESUME

ED 352 920 HE 026 120

AUTHOR Koon, Jeff
TITLE Identifying How To Improve Academic Programs by Linking Various Aspects of Their Delivery to Particular Student Outcomes. ASHE Annual Meeting Paper.
PUB DATE Oct 92
NOTE 43p.; Paper presented at the Annual Meeting of the Association for the Study of Higher Education (Minneapolis, MN, October 29-November 3, 1992).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Outcomes Assessment; *Course Evaluation; Curriculum Development; *Educational Improvement; *Educational Quality; Higher Education; Multiple Regression Analysis; Outcomes of Education; *Program Improvement; Questionnaires; Student Attitudes; Student Educational Objectives; *Undergraduate Study
IDENTIFIERS *Academic Program Evaluation Questionnaire; *ASHE Annual Meeting

ABSTRACT

This paper examines a systematic method of improving undergraduate education by using a variant of the academic program evaluation questionnaire, and offers a model for understanding program dynamics and choosing ways to improve academic programs. Analysis is based on mean ratings from an academic program evaluation questionnaire used to elicit student ratings of various aspects of the delivery of their academic major program and such delivery's contribution to some of their own outcomes. Discussion focuses on the development of the research model and the quantitative analysis used to measure the correlations between outcomes and aspects of program delivery. Findings of the regression analyses address the following areas for the rated outcomes: (1) promoting social responsibility or developing skills useful in the community; (2) providing a broad, general, humanistic, or person-centered education; (3) preparing for a career; (4) preparing for graduate or professional school; (5) training in critical thinking and analysis; and (6) being satisfied to date with the major program. It is noted that over half of the correlations observed between these ratings of value-added outcomes and aspects of program delivery were statistically significant, and many were high in magnitude. (Contains 29 references.) (GLR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED352920

IDENTIFYING HOW TO IMPROVE ACADEMIC PROGRAMS BY LINKING VARIOUS ASPECTS OF THEIR DELIVERY TO PARTICULAR STUDENT OUTCOMES

by Jeff Koon

HE026120

October 30, 1992. Paper presented at the Seventeenth Annual Meeting of the Association for the Study of Higher Education (ASHE), Minneapolis, October 29-November 1, 1992.

(A later version of this paper may or may not in due course be prepared.)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jeff Koon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Texas A&M University
Department of Educational
Administration
College Station, TX 77843
(409) 845-0393

ASSOCIATION FOR THE STUDY OF HIGHER EDUCATION

This paper was presented at the annual meeting of the Association for the Study of Higher Education held at the Marriott City Center, Minneapolis, Minnesota, October 29 - November 1, 1992. This paper was reviewed by ASHE and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC collection of ASHE conference papers.

IDENTIFYING HOW TO IMPROVE ACADEMIC PROGRAMS BY LINKING VARIOUS ASPECTS OF THEIR DELIVERY TO PARTICULAR STUDENT OUTCOMES

by Jeff Koon

Introduction

Although there are numerous exceptions, many faculties and institutions have deflected most of society's demands for accountability. Some legislatures find this frustrating (Ewell, 1991) and more are becoming reluctant to fully fund higher education, especially when state revenues are insufficient to fund other needs. Yet only a few years ago, higher education was a high priority because of its purported value to economic development. Although expectations that higher education could quickly make important new contributions to economic development undoubtedly were too rosy, the most recent fall from grace seems both rapid and far. I would suggest, as some others have also, that an underlying problem is that colleges and universities have not been sufficiently accountable to long-standing demands in the most basic area--undergraduate education--and that the lack of accountability in this area undermines public trust, and contributes to the social impetus toward scandal-making rather than problem-solving in other interfaces between higher education and society (such as billings for research overhead costs, research malpractice and conflict of interest, and sports).

In the most basic realm, society continues to question the quality of graduating students and their educational programs. The "assessment movement" has answered mainly with standardized tests of student knowledge. But such tests yield data about only a few content-related outcomes; and they seldom clearly reveal how best to improve those outcomes, let alone the many outcomes not assessed. Insofar as institutions resist even modest programs to promote institutional self-improvement, or conduct evaluations that lead to little real change, legislators are more likely to become dissatisfied. With the help of "scandalous" stories about unintelligible teaching assistants, the proportion of undergraduate classes taught by teaching assistants and non-regular faculty, and with real shortages of money, it is no wonder that questions have now become more frequent about the amount of work that faculty do, particularly with respect to the teaching of undergraduates (Jacobson, 1992; Winkler, 1992). However, both of these kinds of questions initially surfaced many years ago (e.g., Institute for Research in Social Behavior, 1978), and their current manifestation may be stronger because of the failure to deal with them effectively the first time around.

Whether you prefer this or another sketch of the accountability problem, this paper attempts to address what many students, alumni, legislators, and citizens regard as the basic problem: how to improve undergraduate education. But to use the suggestions offered here, faculty (and administrators) may have to ease up on their paternalism and on their insistence on faculty prerogative, and instead pay more attention to student views. A first suggestion, which is not the subject of this paper, is that teaching be assigned a specific, high weight

in faculty promotion and tenure decisions, and that student evaluations of teaching be used as the measure for awarding the preponderance of that weight in individual cases (Koon, 1986b). The second suggestion is the topic of this paper.

A Systematic Method to Improve Undergraduate Education

For a variety of political and faculty prerogative-related reasons, few campuses have voluntarily chosen to elicit student evaluations of all campus academic programs with a broadly encompassing, systematically analyzed, and periodically administered academic program evaluation questionnaire (APEQ). Nor is the literature in higher education replete with studies pertaining to the reliability, validity, and value of such APEQ-based student ratings. (For example, compare the literature on student evaluations of teaching.) Still, a few authors have reported some evidence about the reliability and validity of student ratings of their college experiences or, occasionally, of departmental programs (e.g., Astin, 1980; Baird, 1976, 1980; Clark, Hartnett and Baird, 1976; Koon, 1986, 1990; Pace, 1985; Pace and Friedlander, 1982; Wise, Hengstler, and Braskamp, 1981). And where such evaluations are elicited, do they not tend to be ignored or resisted when they suggest changes not desired by faculty and administrators, in part because that is typical reaction by any authority (Dressel, 1980; House, 1972)? (Such has been my experience.) And despite some recognition that survey research is among the greatest advances in the social sciences in recent decades (National Academy of Sciences Committee on Basic Research in the Behavioral and Social Sciences, 1982), a disregard for systematically elicited client evaluations is found not only among strict behaviorists (e.g., see Haynes and Wilson, 1979) but among some naturalistic evaluators (e.g., see Guba and Lincoln, 1981). There is great value in the new perspectives on qualitative analysis, but practitioners almost always are members of the professorial class and, consciously or unconsciously, are not aware of assumptions and constituency-balancing habits that serve to protect faculty prerogatives. And when the foxes guard themselves, too much of what is called program review is a rearguard action against change (by which higher education ultimately undermines itself by promoting a reaction--as we have seen, here a mostly reactionary impulse--by society). Moreover, colleges and universities so emphasize their uniqueness that a mandate for systemwide or statewide use of at least some of the most important APEQ items would be anathema to most faculty--even though some degree of common instrumentation across programs would be regarded as almost essential in any other field of research. University faculty and administrators run, but can they hide?

Ah, but the value of APEQs is unproven, you say. Insofar as that is so, faculty in colleges and universities have failed to propose and conduct the relevant research. Moreover, APEQs are relatively inexpensive and item wordings can be formlated to elicit ratings of many outcomes and of almost all aspects of the delivery of an academic program (Koon, 1990). You cannot find a cheaper way to conduct a fairly comprehensive evaluation. Also, because student bodies nowadays are older and more experienced than in earlier decades, their views should be more respected. And campuses probably need to make provisions to analyze student APEQ data by ethnic status.

A typical first level of analysis of student responses to an academic program evaluation questionnaire would compare outcome and effectiveness ratings across programs, perhaps for an entire campus, but ideally, within each disciplinary area (e.g., among social science-related programs). When I began to report our APEQ results to the campus, I thought that student ratings of various aspects of program effectiveness, compared separately for programs within a single disciplinary area, would be sufficient to identify which aspects of program delivery needed improvement. In some respects, they are. But results from a set of ratings convey a sense of priority only insofar as particular aspects of program delivery are rated especially low relative to other programs.

Outcomes, however, are more readily prioritized. What is needed then, is more information about the relationship of the various aspects of program delivery to student outcomes. Thus a program's faculty and students could examine the relative strength of its various student outcomes, and determine priorities with respect to their improvement. In this case, we are talking about outcomes as rated on an APEQ, but nothing prevents the use of other information to help establish their relative need for improvement. Then, estimates of the relative strength of the relationships between the outcomes and the various aspects of program delivery can suggest which aspects of program delivery, when improved, are most likely to yield the greatest improvements in the associated outcomes.

Moreover, insofar as the relationships between particular outcomes and particular aspects of program delivery tend to be consistent across disciplinary areas and settings in higher education, we will be able to learn more about the dynamics of higher education--the cause-effect relationships in student learning. Judging by the modest and rather limited generalizations that Pascarella and Terenzini (1991) were able to make after their massive review, there is much to be learned about educational dynamics in higher education. If nothing else, the relationships suggested by the analysis of this and other APEQs should provide a rich source of hypotheses.

Study Goals

Because this paper represents an early study in this area, my objectives and hypotheses are somewhat exploratory in character. The approach is correlational and the general hypothesis is that the proportion of statistically significant relationships will exceed that expected by chance alone (and, further, that the magnitude of some if not many of the correlations will be high enough to have important implications for program improvement). I should add that such especially positive results were by no means assured and that I undertook this test with some trepidation: what if only 10% of the correlates were significant statistically (at $p < .05$), and what if the strongest correlates ranged only from around .32 to .40 or so? Would that be much more meaningful than we typically find in higher education research?

In any case, the correlations of the mean student ratings of the aspects of the delivery of their academic programs with mean student ratings of their own outcomes also suggest the character of the educational dynamics involved in higher education. Although these dynamics can be thought of as parts of cause-effect relationships, the rated aspects of progam delivery have no clear

hierarchy among themselves (e.g., temporally), so the apparent effects of any given aspect of program delivery, notwithstanding the magnitude of its correlation with a particular outcome, may arise in whole or in part from the effects of one or more other aspects of program delivery.

And if the evidence for the first hypothesis tends to be strong, then I will also need to devise a way, methodologically, to simplify and summarize the educational dynamics observed within this group of programs and, concomitantly, to demonstrate further that, or how, such information might be useful for improving academic programs. This is a second goal of this paper.

I find it useful to think of each aspect of program delivery as a potential **lever** for improving one or more outcomes of the program. This conceptualization helps further to remind us that the levers which might enhance an outcome consist not only of the highest correlates of each outcome, but of any correlates that are meaningful in terms of magnitude. This is important because changes in some aspects of program delivery may be easier to implement than others, may cost less in terms of dollars or effort, may be less likely to induce debilitating conflict within a department, and so on. Moreover, given the comments above about causality, it may be very difficult to induce more than modest improvements in one aspect of program delivery in isolation--because of its relationship with other aspects of program delivery.

This paper offers a potentially generalizable model for understanding program dynamics and choosing ways to improve academic programs. That is, regardless of the findings here, the hope is that this method, or some variant(s) on it, will be usable at almost any campus. In saying this, however, I should add what may be a very important proviso. It is possible that the particular features of our APEQ (e.g., the type of wordings used in the items) were important to developing a questionnaire that effectively distinguished between programs (Koon, 1990, pp. 6:85-92). A major potential problem here is that if data elicited from an APEQ yield no significant differences between programs, the evaluator cannot be sure whether the quality of the programs was in fact nearly equal or whether the questionnaire failed to measure such real differences as existed.

With respect to the data to be analyzed here, I long ago reported comparisons of mean ratings by program, separately by disciplinary area, to the university campus (Koon, 1977a, 1977b, 1979, 1981a, 1981b). I also have reported (Koon, 1986, 1990) on some other properties of our APEQ, including: test-retest reliability; stability of some program mean ratings over time; the differentiating power of the APEQ among programs <u>within</u> each of three disciplinary areas; and a few validity-related indications.

Methods

Our APEQ (the "Undergraduate Major Program Evaluation Questionnaire"), which was subject to ongoing revision from year to year, ultimately included items representing 47 aspects of program delivery that were used in common across 27 or more upper division academic major programs. Of these 47, 35 have data for 38 programs. Also included were 9 outcome ratings obtained for the 27 programs, and 7 for the 38 programs. Item wordings are given in Appendix A.

Each academic program included in the study is represented by the mean rating of respondents from a sample (or all) of its junior and senior undergraduate majors. There are from 21 to 87 respondents per program, 1,387 in all. Ratings were solicited toward the middle of the winter quarter and, absent a response in a few weeks, re-solicited early in the spring quarter. The subjects (academic programs) derive mainly from three samples (Koon, 1977, 1979, 1981a), which included 14 social science and related programs, 12 engineering and physical science programs, and 11 biological science and related programs. Women's Studies (Koon 1981b) was subsequently added to the social sciences group, bringing that total to 15. Sample response rates ranged from 51.4% to 54.2%; program response rates ranged from 37% to 72%.

Nine of the items pertaining to aspects of program delivery, and two items about outcomes, were used only in the later samples, leaving 27 cases and making data unavailable for the biological science programs. These data losses were due mainly to the need to reword or modify items that were not entirely satisfactory in our first APEQ. In addition, some of the revisions to item wordings made in our APEQs were deemed very unlikely to impair comparisons among programs significantly, so the data were used across the entire sample.

For three items (regarding equipment, nondominant schools of thought, and academic-experiential work), "Don't Know/Not Applicable" or blank responses were so prevalent in one or more programs that the mean rating seemed untrustworthy; these items were included in the correlation matrices (with pairwise deletion of missing data by SPSS, 1990), but not in the regression analyses (see below).

For more detail about the samples, questionnaires, cover notes, and so on, see Koon (1990) or the original reports (Koon, 1977, 1979, 1981a, 1981b).

Note: In the abbreviated item descriptions used in the tables and as keywords in the text, a minus (-) as a prefix to the wording indicates that the item is negatively worded with respect to program effectiveness--such that negative correlations with outcomes are expected. (Do not assume that the two minus signs mean that program ineffectiveness is correlated with a positive outcome; but a significant correlation between an outcome and a negatively signed item descriptor should be fairly rare.)

Why Mean Ratings and Not Individual Ratings. Larson (1979) broke down into its sources the variance in student ratings of teachers, identifying which were and were not eliminated in the aggregation of data into mean ratings by course. The situation here is parallel in many respects: The aggregation of ratings into means minimizes nonsystematic error (which is essentially random and therefore tends to wash out in the aggregation) and greatly reduces noncontingent systematic error (such as "classic" response set errors,* which tend to inflate

---

*"For example, if some raters consistently utilize the socially desirable ends of . . . ratings scales (a leniency error), while others consistently stick to the middle of the scales (an error of central tendency), then having both types of raters contribute to the data set might be expected to produce at least a moderate positive correlation between Xobs and Yobs, even in the absence of both actual and assumed covariation between X and Y" (Larson, 1979, p. 205).

correlations when the analysis is of individual ratings or is drawn from two or more measures taken from a single instrument). However, aggregation into means does not eliminate "the raters' average noncontingent systematic response set errors" (Larson, 1979, p. 206). In this case too, there may be more variation by program in response sets than there would be by section within a single course. More importantly presumably, aggregation least effectively minimizes contingent systematic error: The analysis of means minimizes the effects of assumptions idiosyncratic to particular raters but not the effects of assumptions that tend, to varying degrees, to be normative. (These assumptions pertain to the relationship between X and Y, where either is judged based in whole or in part on information about the other and the assumed connections between the two.) Because we are analyzing mean ratings across a variety of academic programs, normative contingent systematic errors may arise not only from students in general (perhaps even as a class phenomenon), but from rater assumptions that differ by disciplinary area, or by program. Such assumption-related differences cannot readily be distinguished from "real" differences in program effectiveness and outcomes.

Control for Disciplinary Area. It is possible, however, to control to some extent for most of the differences by disciplinary area. When used as a covariate, our DiscpArea variable controls for real, response set-related, and assumption-related differences by disciplinary area, i.e., insofar as such differences are linearly arrayed, with Social Sciences = 2, Biological Sciences = 1.5, and Engineering and Physical Sciences = 1. Coded thusly, this control represents, albeit somewhat crudely, a technical/scientific versus human-focused continuum.

To improve our understanding of educational dynamics, variation by disciplinary area "needs" to be controlled only insofar as it is "inherent" to the disciplinary area or based on response set-related or normative contingent systematic error. Ideally, one would prefer to retain any variance that is purely a function of differences in educational effectiveness by disciplinary area, including differences arising from any "unnecessary" adherence to disciplinary area norms.

Despite the imperfections here, when DiscpArea was significantly related to an outcome, I added a second analysis, using it as a control. This enables us to look first at the university's educational dynamics as a whole, that is, at the relationships between rated aspects of program delivery and an outcome while they are affected by disciplinary area, and then to look at educational dynamics insofar as they are consistent across disciplinary areas.

Educational Dynamics and Improving Academic Programs. Program/educational dynamics, which are one key to program improvement, were depicted initially by correlating (Pearson r) the mean student ratings for each aspect of program delivery with each outcome. As suggested above, the first part of the study was simply to examine the magnitudes and significance of the overall correlations. The general hypothesis is that there would be many statistically significant correlations. With 38 subjects, a correlation of .32 is significant at $p < .05$. With 245 correlations tested, approximately 13 (12.25) would be expected to appear to be significant by chance alone (with 2-3 at $p < .01$).

Whatever the strength of the relationships observed here (which represent prospective generalizations about educational dynamics at the upper division level), the results must be regarded as exploratory, pending replications elsewhere. There should be much less to dispute, however, about the value of the methods--a periodically administered APEQ for student majors (and, separately, for students at other levels of their studies), a comparative analysis of program mean ratings within each disciplinary area, and a correlational analysis of the mean ratings across all the programs to better understand local educational dynamics and the "levers" for improving academic programs.

Regression Analyses. Finally, in order to better understand the results with respect to the educational dynamics of the programs and the relationships of each important "lever" (aspect of program delivery) to program improvement, I opted to conduct a regression analysis for each of the seven outcomes for which there were 38 subjects (academic programs), including a variant that controlled for DiscpArea where it had a statistically significant relationship with the outcome measure. In order not to totally ignore the information based on only 27 programs or the three variables for which data were incomplete, the highest correlations of each outcome with these other 12 aspects of program delivery also are noted in the tables presenting the regression analyses.

For these regression analyses to have much meaning with respect to program improvement, however, the reader needs more than a list of the set of variables that maximized the multiple correlation for the mean ratings of an outcome with the mean ratings for 35 items representing various aspects of program delivery, and a table showing the percentage of variance accounted for by each succeeding variable to enter the regression equation. Accordingly, the tables for these analyses show all of the principal zero-order correlations (at $r \geq .40$), and all of the partial correlations above .25 as succeeding variables are entered into the regression equation. The partial correlations represent the correlation remaining between a non-entered variable and the outcome variable after the variation in common with the entered variable(s) has been removed from both. Because a program's faculty and students may choose to implement improvements based on a number of the prospective levers, this approach should be helpful in that it provides some idea about the extent to which a relationship of an aspect of program delivery to an outcome is shared with other aspects of program delivery (particularly for the aspects of delivery that enter the regression equation). Ideally, after viewing data for their own campus in a regression table such as the ones used here, a program's faculty and students might tentatively decide which lever(s) they wish to emphasize in seeking to improve the program, and call their office of institutional research to request a regression analysis with their choice of lever(s) entered first.

The Outcomes. Table 1 presents the intercorrelations among the students' mean ratings of the 9 outcomes and DiscpArea. (Appendix A gives item wordings in full.) Overlap among outcomes (i.e, shared variance) may be based on inherent overlap of concepts (whether in life or as formulated in the items), overlap as perceived normatively by students, and/or overlap due to effects derived from educational practice. For example, we do expect that relatively effective programs will tend to produce better results in several outcome areas, and, consequently, that all the outcomes will tend to be correlated. If these outcomes as measured are to be the focus of our attention, however, it is

preferable that they be at least somewhat independent of each other, lest we be looking at, *de facto*, the same measure over and over.

It should be expected, however, that the students' rating of overall satisfaction with their academic program (abbreviated "AcadSatis") would tend to be correlated with all or almost all of the other outcomes. Such is the case, but two of the eight correlations are not significant statistically.

Further inspection of Table 1 reveals that ratings of whether students will remember for a long time to come what they have learned from a program's courses (abbreviated "Remember") is very closely associated (r= .83) with a program's contribution to understanding the subject matter (CSubKnow). Because data are available for only 27 cases for the latter variable, it was excluded from the series of regression analyses, but the high degree of overlap suggests that the net loss to the study from this exclusion was relatively modest. Similarly it is not very important that a program's contribution to "training in critical thinking and analysis" (CCriticl) was closely related to CSubKnow (r= .67), especially since it is less closely related to Remember (r= 47). Also, a program's contribution to "promoting social responsibility or developing skills useful in the community" (CSocResp) was very closely related to its "providing a broad, general, humanistic, or person-centered education" (CHumanEd; r= .87). Because both of these measures are also highly correlated with DiscpArea, it may be that they would be further differentiated if humanities programs also had been included in the analysis. (As it is, the social sciences tended to be high on both ratings, whereas humanities programs might tend to be higher on CHumanEd and perhaps not as high on CSocResp.). Otherwise (apart from two correlations with the rating of overall satisfaction with the academic program), no two outcome measures share more than 42% of their variance with each other. Perhaps that is about as much independence as we could realistically expect. Finally, as a matter of clarification, the rating of the extent to which the instructional process as a whole has influenced the student (AcadInflu) does not explicitly limit the rating to the influence of the upper division major program (i.e., lower division effects may have been included by some raters); it also has some unusual properties because there are only 27 subjects and because the greater student workload in science and engineering programs almost necessarily yields a somewhat higher average rating of AcadInflu. (Table 3, showing the correlates of DiscpArea other than those shown in Table 1, should be consulted before drawing conclusions about the relationships between AcadInflu and the various aspects of program delivery.)

The Control for Disciplinary Area. As can be seen from the correlations in Table 1, the linear control for the effects of disciplinary area (DiscpArea) was not significantly correlated with program contributions to students' understanding of the subject matter (CSubKnow) or its close counterpart, remembering what was learned (Remember), nor to students' training in critical thinking and analysis (CCriticl). Overall academic satisfaction with the major program also was essentially independent of DiscpArea.

Insofar as these results are replicated in other settings, the independence of these outcomes from DiscpArea would enable fair comparisons among programs across an entire institution, rather than only within a disciplinary area. On the other hand, DiscpArea had a significant correlation with the other five

outcome variables and with some of the aspects of program delivery (again see Table 3). Accordingly, DiscpArea was used as a control to produce a second variant of the regression analyses conducted with these measures.

Results

Correlations Between Outcomes and Aspects of Program Delivery. With respect to the general hypothesis, that there would be observed many significant and meaningful correlations between the mean student ratings of the 7 outcomes and the 35 aspects of program delivery, the results were outstandingly positive, tending to confirm the value of this kind of analysis and of the use of academic program evaluation questionnaires (APEQs) with students. Of the 245 correlations involving 38 programs, 133 (54.3%) were significant at $p < .05$. Only three aspects of program delivery were significantly correlated with six of the outcomes for which 38 programs were rated (and none were with all seven of these outcomes). Table 2 presents these correlational results as well as the data for the other 2 outcomes and the other 12 aspects of program delivery. The magnitudes of the correlations were such that 98 of the 133 significant correlations (or 40% of the 245) had a probability of occurrence by chance alone below .01 ($r = .41+$), and 52 (21.2%) had $p < .001$ ($r = .52+$).

Results and Discussion: The Regression Analyses

Tables 4-10 present the results of the regression analysis for each of the 7 rated outcomes with the 35 rated aspects of program delivery--the variables for which there were 38 subjects/programs. Also noted off to the side or bottom of each table, for the 12 variables with data for fewer than 38 subjects, are any other aspects of program delivery that were highly correlated with the outcome.

In each regression equation, the first variable entered serves, in effect, as the definition of a factor that most strongly seems to affect this outcome. A number of other aspects of program delivery typically are related to that factor (as if they had lower loadings on it). These other aspects may tend in part to be a subset of the first item entered, or they may be independent in essence but often shared in practice by effective programs (and be jointly weak in less effective programs). Per the discussion above, the shared variance also may in part be a function of commonly shared student perceptions as filtered through the questionnaire. However, when the control variable, DiscpArea, is applied, it would eliminate the linear effects of shared group perceptions within the continuum that ranges, as described above, from technical/scientific to human in focus.

Promoting Social Responsibility or Developing Skills Useful in the Community. Data for a "program's contribution" to "promoting social responsibility or developing skills useful in the community" (or CSocResp) will be used to provide a detailed example (see Table 4). First, let us consider some of the statistical events, then we will look at some possible meanings. SocIssues enters the regression equation first, accounting for 66.7% of the variance in the CSocResp outcome, but much of the variance they share is also shared by -IntrUGs ($-.67^2 - .05$ as a maximum = .399+; thus 39.9+% of the total variance in CSocResp also is shared by -IntrUGs at the same time as it is shared by SocIssues). Other variables also sharing large amounts of this same 66.7% of

the variation in CSocResp include: CrseOpts ($.63^2$ - .05/max = 34.7+%), OpnCommcn ($.62^2$ - $.30^2$ = 29.4%), OrientnOK (28.4%), UGonComm (26.9%), and CrseAdvg (26.0%). In addition, a number of other variables also are listed as fairly important zero-order correlates of CSocResp, and should not be ignored or discarded by people interested in greater detail on this subject.

Now consider SocIssues, -IntrUGs, and CrseOpts together. When faculties give more attention to social issues in teaching their subject matter (SocIssues--the lever apparently most likely to improve this outcome), they may also seem to be more interested in or concerned about undergraduates, or their willingness/effort to give more attention to social issues may be a reflection of a greater interest in or concern about their undergraduates, but the program does not thereby of necessity provide more curricular course plan options with which students can pursue differing, somewhat more specialized interests. The latter is a practice, however, at least in this sample, that tends fairly strongly to accompany greater attention to social issues, perhaps because it implicitly opens a dialog about pursuing one or another of the specialized options, and the reasons therefor, some of which will be associated with social issues. (The reasoning here may be more obvious to the reader if the choices are between a behavioral and humanistic focus in psychology than between organic and inorganic chemistry.) Differing options by which to major also may be harder to arrange in fields in which knowledge is highly cumulative or course requirements subject to accreditation. This discussion suggests the kinds of thinking that also might be applied to the other variables listed in the first paragraph above.

After the first step in the regression equation has been completed: Of the variance remaining in CSocResp (33.3% of the original total), 36% (r^2= $.60^2$ = .36) is shared by CareerInf, a rating about whether the program and faculty provide enough information about prospective careers for those with a bachelor's degree in the field, resulting in an increase in the multiple correlation (R^2) of .119 (.333 X .36 = 11.99%) when CareerInf is entered into the regression equation. That part of the variance in CareerInf that was not overlapped with the first item ("factor") that was entered into the regression equation, but that is correlated with the remaining variation in CSocResp, becomes the second "factor" to enter the regression equation. (The correlation of CareerInf with CSocResp rose from .45 to .60 after SocIssues was entered into the equation because the .60, a partial correlation, represents the proportion of the remaining variance they share after the effects that both jointly share with SocIssues has been set aside. At the outset, the zero-order correlation of SocIssues and CareerInf was .06, so very little of the variance in CareerInf is likely to have entered with the first item/factor.) Many fewer items mutually share the remaining variation between CSocResp and CareerInf, but OrientnOK has a "loading" on this item/factor too ($.57^2$ - $.38^2$ = 18.1%). Apparently a good, general orientation meeting for students new to the major involves discussion and questions about possible course selections (even if there is only one option and a few electives) and career paths, including specific social issues associated with curricular plan options, and individual courses and careers. This interpretation also is suggested by the variance in factor one that is shared with course advising and open communication between faculty and students. However, the career-related variation in OrientnOK was mostly excluded from the first factor and included in the second. These results further clarify that the

first factor's link to "promoting social responsibility and developing skills useful in the community" tended to exclude the outcome aspect pertaining to "developing skills useful in the community," whereas the second item/factor focused on career-related development.

In the next step in the regression equation for CSocResp, UGonComm, or student assent with the statement that "the program should (even if it already does) include undergraduates on most committees, provide an office, allow a quarterly mailing, etc., to facilitate undergraduate organizing and input," increases R^2 to .820, accounting for 16.0% of the remaining variance in CSocResp (which was after the second step reduced to 21.4% of the total initially), or 3.4% of the original, total variance (as shown by the increase in the multiple R^2). The item expresses student support for such student participation but is ambiguous with respect to the performance of the programs. However, the responses seem to reflect positively on programs that support student participation, in that fewer students feel any need to accept faculty views in opposition to student participation in the department. (In some engineering programs, student involvement in pre-professional societies is an important curricular adjunct for facilitating contacts to practitioners and the subsequent completion of a requirement for experience in the field before the graduate can be officially registered as an engineer, and many faculty members were supportive of these student organizations.) For students actually involved in undergraduate organizations, of course, there can be an important contribution to leadership skills, learning about issues in the department/field, and so on. (Remember that much of the variance that UGonComm originally shared with CSocResp was also shared with SocIssues and, hence, was incorporated into the first factor to enter the regression equation.)

Toward the end of the regression equation, OfcStaff, the extent to which program office staff are helpful (possibly in programs in which career-related information and orientation are insufficient) increases the multiple R^2 to .855. And -LDPreReq, the extent of student agreement that some of the lower division prerequisites are not really necessary for the upper division program, increases the multiple R^2 to .872 (R= .934). Thus, some of the "unnecessary" prerequisites may make a very small contribution to CSocResp (possibly through the development of skills useful in the community). Or, because this item is among the last to enter the regression equation, shared error variance between it and CSocResp could have been paired up.

The Supplementary Items About Aspects of Program Delivery. In addition, some supplemental correlations, based on fewer programs/subjects, are reported. Among these are NonDomToo--an item about whether there are enough courses that explore the contributions of nondominant schools of thought in the field--which is very highly correlated with SocIssues (r= .81) and which, therefore, though it casts a somewhat different light on the question about social issues, presumably would have been largely redundant in the regression equations. The correlation of SocIssues with FacTeachg--an item about whether program faculty teaching is "challenging, of high quality, and increases my interest in the subject matter"--was .55.

The items about HavFacAdv--"I have a regular faculty advisor (or otherwise know a faculty member) to whom I would feel comfortable going at any time during the year"--and FcCntctSf--"I have been satisfied with the level of interaction I have had with professors here"--were correlated with CSocResp at .54 and .47, respectively. These items, along with FcDiscsns (a self-report measure of out-of-class discussions with faculty members rather than a rating), represent some of the more informal aspects of student-faculty contact. Fortunately, these three variables also are fairly well represented in the 38-program analyses by variables such as WriteRecn (r= .68, .81, and .78, respectively), OpnCommcn (r= .91, .86, and .60), and EnufSmCls (r= .71, .72, and .70), as well as several others correlated at only slightly lower levels (e.g., -IntrUGs, Feedback, -Explain).

The percentage of T.A.s judged as effective teachers, abbreviated TAEffctvT, also was related to CSocResp. Unfortunately, there are few good substitutes for this measure, the best being GradInfo (r= .60), -Explain (r= -.56), and the two methodology items, -Probsolv and -TradMeth (r= -.56 and -.53, respectively). (These correlates do, however, suggest things about the role of T.A.s in an upper division context and seem to have some face validity.)

Lastly, EnufFldWk--"There are enough fieldwork, internship, group, lab, or other academic-experiential course offerings in this program, and/or credit is readily available for appropriate types of volunteer experience"--was correlated more with CSocResp (r= .47) than with any other outcome. There are few viable substitutes for EnufFldWk among the 38-program variables; though OrientnOK is correlated with it at r= .58 and OpnCommcn is correlated at r= .57, both are much more strongly related to many other variables.

The Control for Disciplinary Area. The second summary at the bottom of Table 4 shows what happens to the regression equation for CSocResp if DiscpArea, the control for disciplinary area (representing a technical/scientific versus human-focus continuum), is forced to enter the equation first.

In this case, the partial correlations after DiscpArea is first entered into the equation also are of interest. DiscpArea, with a zero-order correlation of .59 with CSocResp, accounts for 35.1% of the variation in that outcome, leaving 64.9% remaining. Although I have not provided a separate table showing the regression equation for CSocResp with DiscpArea entered first, the results are somewhat similar to those shown in Table 5A. (Table 3 also provides a rough idea of the impact of DiscpArea on the mean ratings of the various aspects of program delivery.)

Providing a broad, general, humanistic, or person-centered education. There may be some risk of boring the reader by reviewing the most similar results next, but the similarity also enables the review to be short, except insofar as major differences in the analysis are important, or minor differences are worth highlighting.

Because the SocIssues variable is the first to enter the regression equation for both outcomes, in comparing the results for program contributions to "providing a broad, general, humanistic, or person-centered education" (CHumanEd) with "promoting social responsibility or developing skills useful in the

community" (CSocResp), we might first note which zero-order correlations most differed in magnitude. Viewed in terms of the aspects of program delivery which most contribute to them, the chief differences between CHumanEd and CSocResp were: requiring too many courses for the major had a more negative effect on the CHumanEd outcome (a difference of 27% in the variance shared between the outcomes and -ReqCrses, $-.62^2$ vs. $-.34^2$, respectively). Similarly, demanding a lot of work for the number of units awarded was more of a deterrent or detriment to CHumanEd (a difference of 24%); the control variable, DiscpArea--focusing on technical/scientific disciplines as contrasted with human-oriented disciplines--also had a larger negative effect on program contributions to CHumanEd than to CSocResp (DiscpArea, a difference of 20%); providing enough information about careers was less important to the CHumanEd outcome than to CSocResp (a 16% difference); and grading very rigorously also is contraindicated more for the CHumanEd outcome (a 14% difference), perhaps in accord with parts of the Pass/NotPass grading philosophy.

As can be seen by comparing Table 5 and Table 5A, which presents the full results for the regression on CHumanEd with DiscpArea entered first, some of the largest differences between technical/scientific fields and the social sciences account for many of the results shown in Table 5. These differences also are evident in the mean ratings by disciplinary area (see Appendix A). Specifically, there is an almost total lack of electives outside the upper division engineering programs (and often, in effect, in the pre-med programs as well), and heavier workloads and more stringent grading curves typically are observed in the sciences. Thus, when we look at the results for the two regression equations, they differ considerably.

At the university-wide level then (Table 5), the programs contributed most to the CHumanEd outcome: if faculty show enough concern about social issues in their teaching (SocIssues); if program students actually received a call to a meeting of their program's undergraduate organization (-NtcUGMtg), if students support (and/or faculty support/allow) student participation on program committees and student organizing generally (UGonComm), and if students urge that faculty use the rating form from the campus student government's professor/course guide to evaluate their large or required courses (TEvalGuid); if workload per unit is not unduly high (Wrkload/U); and if the program's system of advising "makes it easy for me to choose courses, professors, and to understand how best to meet my academic needs" (CrseAdvg). Overall, the regression equation on CHumanEd accounted for 90.9% of its variance ($r = .953$).

Once disciplinary area (DiscpArea) is controlled (Table 5A), we can get a better idea of what all programs can do to improve, largely irrespective of their disciplinary area-related character and constraints. As suggested by the order of entry in Table 5A, "a broad, general, humanistic, or person-centered education" is most facilitated in any of these disciplines by a good, general orientation for students new to the major (OrientnOK), the absence of too many required courses (-ReqCrses), tests that the students regard as accurate measures of what they have learned (TestAccu), methodology courses that don't provide "enough experience in real problem-solving and analysis" (-ProbSolv, here entered as a plus), student support for student participation in program decision-making and student organizing, which sometimes needs faculty support as a facilitator (UGonComm), and a generally helpful office staff (OfcStaff). Including the

variance accounted for by the control variable (DiscpArea), overall, the regression equation on CHumanEd accounted for 92.3% of its variance (r= .961). The reversal on -ProbSolv suggests that repeated in-depth problem-solving (such as in a laboratory course for prospective majors in a basic science course) may not be as useful for breadth as, say, a "physics for poets"-approach.

Of course the evaluator should also search for other levers by which to improve CHumanEd by examining the partial correlations in the regression equation, especially just after the control variable has been entered, to ascertain, for example at step 2, the variables that have partial correlates nearly as high as OrientnOK, and that may in part substitute for it, or be a second but also necessary focus of effort if substantial program improvement in this area is desired.

Preparation for a Career. Because the regression equations were basically the same with or without the forced entry of DiscpArea as a control, the regression on program contribution to "preparation for a career" (CCarPrep) can be described relatively easily. As can be seen from Table 6, DiscpArea was the second variable to enter the regression equation when it was not forced to enter first; and the first variable to enter in the unforced, stepwise regression, CareerInf, was second when DiscpArea was forced in first.

So, whether the focus is university-wide or within the disciplines represented in this sample (i.e., remembering that the control is both linear and somewhat limited definitionally), there were four aspects of program delivery that contributed significantly and uniquely to CCarPrep. First was "the faculty and program provide enough information about career opportunities for students with a bachelor's degree in this major" (CareerInf). Because the engineering programs, even if not the sciences, are more directly tied to a career, disciplinary area (DiscpArea) is another factor. (Classifying the business major with the social sciences, and geology, physics, chemistry and math with engineering, presumably reduced this effect.) The presence of too many required courses for the major (-ReqCrses), as rated by the same students, also contributed to CCarPrep. (This result may arise in part because DiscpArea does not control for all of the differences between disciplinary areas, and/or because program requirements may in fact make an important contribution to CCarPrep yet be "too many" from almost any other perspective.) Finally, insofar as a program's courses were scheduled at times that delayed, made it impossible, or made if difficult for their student majors to actually enroll in them (-Schedule), there apparently was a negative effect on CCarPrep. Overall, the regression on CCarPrep accounted for 85.8% of its total variance, which represents a correlation of .926.

In addition, during the course of the calculation of the regression equation for CCarPrep, one item entered the equation and was subsequently removed from it. This item is "I trust the faculty to take student needs and views into account in their decision-making for the program" (TrustFac). Although it added substantially to the equation when it first entered, the addition of subsequent variables reduced the unique contribution of TrustFac so much that its removal from the equation improved the reliability of the results at little loss to the magnitude of the multiple correlation between the outcome and a combination of the aspects of program delivery.

Of course the evaluator should also check the aspects of delivery that are correlated with CCarPrep but do not enter into the regression equation. In this case, at least prior to the entry of DiscpArea, some zero-order correlations may be misleading from a perspective within any given disciplinary area (if not quite so much so from the university-wide perspective). In particular, in the engineering and sciences programs, it is unlikely that increasing the workload per unit (Wrkload/U) or adding to the number of courses required for the major -ReqCrses) would further contribute to CCarPrep, and the same may be true of the social sciences. Because links to specific careers, course requirements tied to accreditation, and subject matters that are highly cumulative are not readily affected by programs at any given college or university, it seems likely that the preferred solution for the regression on CCarPrep is the one in which DiscpArea is forced in first. But otherwise, having enough ethnic minority faculty members (EnufEthnF), at least in programs in which many ethnic minority students are enrolled, having able and motivated student peers in the program (AbilMotvn), being able to study and learn effectively at whatever level of competition among students is present (Competitn), and having advanced courses available in all of the students' interest areas (AdvnCrses), might serve generally as important levers for program improvement with respect to preparing students for a career.

Preparation for Graduate or Professional School. In the case of program contributions to "preparation for graduate or professional school" (CGradSch), DiscpArea was the highest correlate, so it entered first in the stepwise regression equation. (See Table 7.) I have no information on the reality here, that is, whether the engineering and science students actually were more likely than the social science students to go on to graduate or professional school (in general, they aren't, per Pascarella and Terenzini, 1991). But there are many factors that could have shaped these responses, including the relative effectiveness of the programs (especially in the social sciences, in which there was a very wide distribution of mean ratings on most items), market demand for people with baccalaureate degrees and for more advanced professionals in various fields at the times of data collection, the presence of many pre-meds in the biological sciences (many of whom probably would be disappointed when they actually applied), and the small cadre of undergraduate majors in programs such as math, physics, and geology (fields with weak ties to careers at the baccalaureate level but very few undergraduate majors compared to the number of faculty members). Because program faculties can do very little about many factors affecting whether students continue on to graduate or professional school, it again appears that the most meaningful analysis should begin with the control for DiscpArea and that the search for additional levers for program improvement should normally discount aspects of program delivery that have high zero-order correlations with CGradSch if their partial correlation declines markedly when DiscpArea is entered into the regression equation. Also, program faculty generally were interested primarily in research and graduate students; as a result, another confounding factor is that the undergraduate student majors in some programs, on the average, have even been known to rate the "importance to you" of "preparation for graduate or profesional school" lower than their program's contribution to that goal.

Apart from DiscpArea then, five other items were included in the regression equation for student ratings of their program's contribution to preparation for graduate or professional school. These were whether the student had needed

further explanation or extra instruction in his/her subject matter "but help or a referral did not seem easy to obtain" (-Explain); whether there were advanced courses in the program, or otherwise available, in all his/her areas of interest (AdvnCrses); the absence of enough course plan options by which to pursue the major (CrseOpts); less reliable receipt of notices about a general organizational meeting of undergraduates in the student's program (-NtcUGMtg); and less felt need for faculty to solicit student evaluations of teaching with the rating form from the student professor/course guide (TEvalGuid). Together with DiscpArea, these variables accounted for 79.8% of the variation in CGradSch, reflecting a multiple correlation of .893. (When only a single course plan option is available, it most assuredly will provide training suitable for graduate school, whereas multiple options may include a second career track, a second liberal arts focus, etc.)

Among the other aspects of program delivery that were correlated with CGradSch but had their shared variance subsumed by the listed items above (i.e., discounting those subsumed by DiscpArea) were a trust that faculty will take student needs and views into account in their decision-making for the program (TrustFac), the rated ability ("caliber") and motivation of the student's peers (AbilMotvn), and a denial that methodology courses did not provide enough experience in real problem-solving and analysis (-ProbSolv). In addition, among the variables with fewer than 38 subjects/programs, two were significantly related to CGradSch and were not significantly related to DiscpArea, such that they should be considered as levers for program improvement too. These are the self-report of "about how many times did you have a discussion outside of the classroom with any professor, apart from study-list advising, in the last three months of instruction (i.e., excluding vacation periods)" (FcDiscsns) and the extent to which the coursework is interesting and challenging (ChalIntrg). Based on the generalizations in Pascarella and Terenzini (1991), and face validity, most of the correlates of CGradSch in the paragraphs above seem fairly appropriate, though CrseOpts, -NtcUGMtg and TEvalGuid weren't expectable.

Training in Critical Thinking and Analysis. We have now completed treatment of the outcomes that needed DiscpArea as a control. The results pertaining to a program's contribution to "providing training in critical thinking and analysis" (CCriticl) were both the simplest and least informative of any obtained. See Table 8. The regression of the various aspects of program delivery on CCriticl accounted for only 45.3% of its variance, reflecting a correlation of .673 between CCriticl and the one variable entered into the equation. The aspect of program delivery that best predicted CCriticl was "the curriculum is structured so that all majors will have a sufficient number of small classes or seminars" (EnufSmCls).

Although this makes a lot of sense (see Pascarella and Terenzini, 1991), because only one variable was entered, it is all the more important to review the zero-order correlations for other possible levers for program improvement. In addition, several of the ratings with fewer than 38 subjects/programs are of interest. However, with the help of Table 8, this is such a straightforward task that there is little reason to guide the reader through it. Suffice it to say that, among the other correlates of CCriticl, there is a fairly high concentration of variables pertaining to faculty-student interactions and involvement, a finding also in accord with Pascarella and Terenzini (1991).

Remembering What Was Learned. "I will remember a lot of what I have learned in the program's courses for a long time to come" (Remember) is our item about subject matter retention. The regression equation for Remember accounted for 81.1% of its total variance (R= .901). See Table 9. Among the more interesting phenomena in this equation is that the first variable entered, "Most of the faculty in this program are receptive to student input on class structure and content" (StuInput), was subsequently removed from the equation because, after three other variables were entered, it no longer shared a significant amount of unique variance with Remember. If nothing else, high ratings on StuInput presumably imply faculty willingness to follow-up in greater detail on topics on which students have questions or need further instruction, and entertain other student ideas about the course. That the highest correlate was in this case removed again suggests the importance of not relying solely on the items included in the regression equation in the search for ways to improve academic programs.

Apparently more essential in the regression equation on students' remembering what they had learned for a long time to come were: a curriculum in which courses related well to each other (-FrgmtCur), enough faculty attention given to social issues in the classroom teaching (SocIssues), tests in most courses that were fairly accurate measures of what the student had learned (TestAccur), peers that were of high caliber or highly motivated (AbilMotvn), and the provision by faculty and program of enough information about career opportunities for students with a bachelor's degree in the field (CareerInf). Given the links of these aspects of program delivery to known ways to facilitate learning--an integrated curriculum/program that hooks into student movitation through social issues and career information, accurate assessment and feedback, and motivated peers with whom to interact, these results seem to have considerable face validity if not also construct validity based on previous research.

As suggested in the list of additional correlates (i.e., those based on fewer than 38 subjects), the vitality of faculty teaching and the challenge/interest of the coursework (FacTeachg and ChalIntrg) are perhaps especially important levers to promoting student retention of what they have learned. But a number of other items pertaining to formal and informal student contact with faculty also appear among the highest zero-order correlates in both columns (OpnCommcn, -IntrUGs, WriteRecn, SfFcCntct, FcDiscsns, HavFacAdv). The correlates of Remember also seem to include a high proportion of items more specifically associated with program content: the ready availability of further explanations of subject matter when needed (-Explain); sufficiently thorough feedback on all aspects of student work (Feedback); enough experience in real problem solving and analysis (-ProbSolv) with up-to-date methodologies (-TradMeth); enough courses that explore nondominant schools of thought in the field (NonDomToo, which is closely associated with SocIssues); and the absence of too many required courses for the major (too many stimulates boredom, memorization and forgetting). Logically enough, the effectiveness of teaching assistants (TAEffctvT) and enough small classes in the curricular structure (EnufSmCls) seem to play a facilitative role here. Further, remembering what one has learned is better enabled when advising has helped the student to choose courses that best meet his/her academic needs (CrseAdvg) and a good general orientation for students new to the major has been provided (OrientnOK). Then

add the regular use of teaching evaluations to keep faculty apprised of the effectiveness of their performance as instructors (TEvalUse) and some student participation in the program (UGonComm); these, together with a sound and responsive program and good student-faculty relationships (as above) also produce more trust among students that faculty will take student needs and views into account in their decisions for the program (TrustFac). And information should be provided about graduate and professional school to which the students might wish to apply (GradInfo). At our university, these levers produced undergraduate major programs in which students believed they will long remember much of what they have learned.

Satisfaction To Date with Your Major Program. Overall satisfaction with the academic major (including "faculty, teaching, advising, courses/curriculum, testing-grading, governance, etc.") is "a fairly comprehensive surrogate measure [of the value-added by a program] that, unfortunately, also may include some level-of-quality effects" (Koon, 1991, p. 6). Thus our item about overall satisfaction with the major program (AcadSatis) presumably is a composite of the outcomes considered previously as well as of other outcomes, hopefully mainly of lesser importance, that were omitted from our questionnaire or included only indirectly. Because of its composite character, we might expect that more of the various aspects of program delivery would be significantly related to it, that more would enter the regression equation, and probably that more of the variance in AcadSatis would be accounted for. As it turned out, AcadSatis had as many (in one case) or more correlates over .40 than did any other of our outcomes; it took 9 stepwise entries before there were no more aspects of program delivery that could make a significant contribution to the regression equation; and 94.4% of the total variance in AcadSatis was accounted for (R= .972)! See Table 10.

What is the simplest path to overall satisfaction with the academic major program, as described by the regression on AcadSatis. AcadSatis was facilitated by open and fairly extensive communication between faculty and students (OpnCommcn); high caliber and/or highly motivated peers (AbilMotvn); grading that was seen as not unduly rigorous (GradeVRig); coursework in which faculty do not state clearly at the outset what students will need to know or do to get an "A" (A=FcState); a level of competition that best helps the student to study and learn (Competitn); courses that relate well to one another, enabling the student to get an integral perspective on the field (-FrgmtCur); sufficient attention to social issues by faculty in the classroom (SocIssues); insufficient program or faculty-provided information about possible graduate or professional schools (GradInfo); and not having enough women faculty in the program (EnufWmnFc).

We recognize some of these variables from the regression equations for Remember (-FrgmtCur, SocIssues, and AbilMotvn), and for CSocResp and CHumanEd (SocIssues). Further, OpnCommcn was highly correlated with these three outcomes as well as with CCriticl (especially).

The negative link between rigorous grading practices (GradeVRig) and overall satisfaction is of special interest. The item, which was intended to convey the idea of too much "rigor," reads: "Grading in this program is more rigorous than in most--it is very hard to get an 'A'." Is the correlation with overall satisfaction a matter of students dissatisfaction with grading on the curve or unduly strict (quasi-)"absolute" standards? Or, as others would have

it, is it a case of the erosion of standards and the desire for an easy "A" that comes into play here? Because peer caliber and motivation (AbilMotvn) had little or no relationship with GradeVRig (r= .15), its entry as the previous step in the equation probably did not so much reveal a hidden relationship between GradeVRig and AcadSatis as simply allow for the original amount of variance shared between GradeVRig and AcadSatis to increase as a proportion of the decreased variance remaining in AcadSatis. The best clue I have thus far uncovered pertains to the differential correlations between GradeVRig and AcadSatis by disciplinary area. The anomaly is in the biological sciences programs, the middle group in the DiscpArea coding, so there is no chance of finding it in the correlation between DiscpArea and AcadSatis. As observed (and the numbers of subjects/programs are too few to make the data trustworthy, but that wouldn't prevent it from finding its way into the later steps of a regression equation), the correlation between AcadSatis and GradeVRig is essentially zero for the other two samples, but is -.69 for the biological science programs. Although grading rigor in the biological science programs was rated the same, on the average, as grading rigor in the engineering and physical science programs, the heavy competition for grades in those biological science programs in which a sizable proportion of majors were pre-meds (and the accompanying lower degree of interpersonal support among peers), may have produced a less satisfactory experience--not only perhaps for the pre-meds but especially for those in more liberal-arts oriented course plan options in these programs (who took many courses with the pre-med students). At the time the data were collected, competition for slots in medical schools was very high, grade-point-average and MCAT scores were of great importance in admissions decisions, the pressure for pre-meds to cheat may have been especially high, and rumors were occasionally heard of pre-meds sabotaging others' lab experiments. Moreover, the rigor of faculty grading was regarded by some students as a deliberate faculty effort to screen students for medical school (but, insofar as undue rigor existed, it may have been more associated with high enrollments and large, crowded lecture courses). In any case, many students felt that they had to make a dissatisfying choice to sacrifice some other educational opportunities/values in the collegiate experience in order, given their priorities, to earn good grades.

A separate item on competition (Competitn) also entered the regression equation for AcadSatis. Although the correlations between AcadSatis and Competitn were virtually the same when calculated separately by disciplinary area, in terms of the mean ratings by disciplinary area, students in engineering and physical science programs were more likely to agree that they studied and learned best with the level of competition which they actually encountered (3.11 vs. 2.67 and 2.71 for the other two samples, E-probability < .01). The average rating here is a little below the middle of the 1-5 scale, which does not suggest a close fit between students and institutional practices but corresponds well with ongoing student complaints about the level of competition at the university.

Another variable that entered the regression equation for AcadSatis may be tied less to competition than to the large, impersonal, alienating university environment described by many undergraduate students. The item, A=FcState, is worded as follows: "Faculty in my program clearly state at the outset what the courses will cover and what students need to know or do to get an 'A'." The equation suggests that affirmative responses (by program) to this item are associated with more dissatisfaction, and vice-versa. Perhaps satisfaction

arises more readily from an educational ambiance in which faculty members do not need to specify in advance the full content of a course (e.g., in order to be fair), thereby constricting course content and student initiative/exploration. In other words, satisfaction may be enhanced by more small classes in which students can discuss, question and reflect.

Also entering the regression equation for AcadSatis was a negative relationship between satisfaction and whether the faculty and program provide enough information about graduate and professional schools to which their students might wish to apply (GradInfo). This relationship accounted for only about 1.5% of the variation in AcadSatis and was in a direction opposite to the zero-order correlation; because it is in an unexpected direction on the eighth step in a regression equation, the observed relationship may be capitalizing on random error. If this variation is meaningful, perhaps the great emphasis on research and graduate education at the university induces so much unexpected student interest in attending graduate school that student needs for information cannot readily be met by faculty offering effective programs; or, perhaps the amount of emphasis on research and graduate education already more than suffices in programs in which career-oriented students are relatively more numerous, such that a small amount of AcadSatis is associated with more faculty acceptance of (and empathy for) students who instead plan to go directly into baccalaureate-level careers. (See also the first paragraph about CGradSch.)

Finally, whether there are enough women faculty in the program (EnufWmnFc) also included a little negative variance related to AcadSatis (1% of the total for AcadSatis), suggesting that some programs' students were disproportionately satisfied academically relative to the shortage of women faculty members in their program. Since this is somewhat anomalous directionally, I wonder if it could be attributed to differences by program in the proportion of students who resent affirmative action and showed it, consciously or unconsciously, on our APEQ. Or, at the ninth step in the equation, it could easily be due to random error.

Finally, and in many respects more importantly than the later steps in the regression analysis, there also are many other ratings of various aspects of program delivery that correlated initially with overall satisfaction with the academic major but which did not enter the regression equation. These correlates could be among the most important levers with which to improve student satisfaction. Included, and probably foremost, are the quality of teaching and coursework (see the list of "other" items in Table 10). The relevant variance of many other variables was mostly subsumed by the entry of OpnCommcn. The items involved include a variety of variables pertaining to student-faculty relationships, both formal--such as course-related advising and orientation--and informal--such as out-of-class discussions and writing letters of recommendation. And several course and curriculum-related aspects of program delivery, such as whether there were enough small classes, the adequacy of course scheduling, the availability of further explanatory help or instruction when needed, accurate testing, thorough feedback on student work, methods courses that provide enough experience in real problem-solving and analysis, and, at least when appropriate, different course plan options by which to pursue the major. Also relevant are student trust that faculty will take student needs and views into account in their decision-making for the program, faculty use of student evaluation of teaching forms, and some degree of student organizing. Completing the list are

teaching assistants who are effective teachers, and the provision of enough information about careers for students with a bachelor's degree from the program.

Implications and Conclusion

It seems to me that this method of analysis has more than demonstrated its utility as a way of identifying how best to improve undergraduate academic programs. Further, the kind of analysis undertaken here, when replicated in many other settings, may well reveal much more than is currently known about cause-effect relationships in higher education, especially with respect to the magnitude of effects. If nothing else, the methods used here can be used to generate a variety of highly meaningful hypotheses about the effect of various aspects of program delivery on various student outcomes.

The analysis was based on mean ratings from an academic program evaluation questionnaire (APEQ) used to elicit student ratings of various aspects of the delivery of their academic major program and its contributions to some of their own outcomes. Over half of the correlations observed between these ratings of value-added outcomes (mostly) and the aspects of program delivery were significant statistically, and many were high in magnitude. The form of presentation chosen here was used to emphasize that the initial correlations between the aspects of program delivery are the primary menu of choices to provide information about how to improve the program. However, because the regression equations reveal some of the joint sharing of variance between two or more aspects of program delivery and an outcome (see the partial correlations), they can help users to identify and select more appropriate combinations of aspects of program delivery with which to produce improvements in one or more outcomes. For example, sometimes, especially in the early steps of a regression equation, the uniqueness of the underlying "item/factors" associated with an outcome seemed important. Moreover, at the local level, the regression equations could easily be recalculated based on tentative department plans for improving an outcome (or perhaps several outcomes jointly!). In addition, the regression equations show the extent to which variation in the outcome (the multiple R^2) was apparently predicated on the aspects of program delivery. The equations calculated here accounted for over 90% of the variation in one outcome and in one surrogate outcome, over 85% in two other outcomes, 80% in one outcome and almost 80% in yet another, and 45% in a seventh outcome.

Although the magnitude of the co-relations between the various aspects of program delivery and the outcomes rated here may well not be highly generalizable, the directionality of almost all of the stronger relationships seems likely to me to be stable (judging by what evidence I could find in Pascarella and Terenzini, 1991). I would expect, of course, that the replicability of the magnitude of particular correlations would tend to decline insofar as the setting differed from the one studied here, a major research university. But a meta-analysis across 20-40 subsequent studies from different settings might provide us with much better estimates of the magnitudes of cause-effect relationships in higher education.

Table 1. Correlations of Mean Ratings of 9 Student/Program Outcome Variables Among Themselves and with Disciplinary Area (as coded**)
(N= 38 academic programs, except *)

Variable	Acad Influ*	CSubK*	CCrit	CHumn	CSocR	CGrad	CCarG	Acad Satis	Discp Area
Remember	.17	.83c	.47b	.59c	.54c	.23	-.02	.77c	.27
AcadInflu*		.41a	.29	-.34	-.21	.65c	.50b	.24	-.61b
CSubKnow*			.67c	.45a	.46a	.44a	.27	.82c	.08
CCriticl				.42b	.47b	.28	.34a	.56c	.02
CHumanEd					.87c	-.38a	-.22	.56c	.74c
CSocResp						-.26	.13	.56c	.59c
CGradSch							.44b	.38a	-.60c
CCarGoal								.29	-.45b
AcadSatis									.15

*Correlations reported for Acad Influ and CSubKnow are all based on mean ratings for 27 subjects/programs (i.e., exclude the Biological Sciences sample).

**Disciplinary Area is coded as follows: 1=Engineering and Physical Science Programs; 1.5= Biological Science Programs; 2= Social Science-Related Programs.

a= Significant at $p < .05$; b= significant at $p < .01$; c= significant at $p < .001$. Two-tailed tests were used.

Table 2. Correlations Between Mean Student Ratings of Aspects of Program Delivery and Aspects of Outcome (N= 38 or 27* upper division major programs)

Aspects of Program Delivery	Outcomes								
	Remember	AcadInflu*	CSubKnow*	CCritic1	CHumanED	CSocResp	CGradSch	CCarPrep	AcadSatis
-LDPrereq	-.26	-.01	-.36	-.25	-.20	.01	-.12	.27	-.27
-ReqCrses	-.47b	.47a	-.27	-.22	-.62c	-.34a	.30	.53c	-.27
CrseOpts	.38a	-.28	.50b	.25	.57c	.63c	-.25	.02	.42b
-FrgmtCur	-.65c	-.16	-.68c	-.33a	-.31	-.26	-.35a	-.16	-.70c
LgLectrOK	-.26	.22	-.22	-.32a	-.32a	-.34a	-.05	-.01	-.25
EnufSmCls	.48b	.15	.60b	.67c	.41b	.43b	.31	.25	.63c
-Schedule	-.39a	-.31	-.62b	-.46b	-.14	-.23	-.37a	-.38a	-.54c
AdvnCrses	.25	.37	.53b	.28	-.18	.02	.55c	.46b	.39a
NonDomToo(N=29)	.73c	-.29(N=25)	.46a(N=25)		.65c	.81c	.79c	-.17	-.21	.58b
EnufFldWk(N=37)	.33	-.15(N=26)	.29(N=26)		.13	.40a	.47b	-.11	.07	.35a
-ProbSolv	-.58c	-.25	-.67c	-.43b	-.11	-.26	-.45b	-.29	-.52b
-TradMeth	-.55c	-.27	-.74c	-.29	-.12	-.16	-.30	-.16	-.36a
-EquipEtc(N=22)	.03	.09(N=21)	.09(N=21)		.18	.12	.14	-.13	.19	.01
ChalIntrg*	.86c	.42a	.84c	.63c	.36	.35	.51b	.22	.83c
Feedback	.65c	.13	.59b	.39a	.44b	.43b	.15	-.05	.51b
-Explain	-.69c	-.02	-.74c	-.26	-.50b	-.43b	-.18	.11	-.59c
SocIssues	.67c	-.40a	.48a	.46a	.82c	.80c	.36a	.25	.65c
FacTeachg*	.86c	.27	.78c	.60b	.55b	.54b	.29	.09	.86c
TAKnowlbl*	.29	.16	.33	.19	.07	.10	.25	-.10	.31
TAEffctvT*	.54b	-.05	.44a	.40a	.44a	.49a	-.01	-.05	.50b
SfFcCntct*	.63c	.05	.63c	.50b	.45a	.47a	.24	.16	.75c
FcDiscsns*	.61b	.47a	.70c	.60b	.07	.06	.64c	.25	.61b
WriteRecn	.65c	.20	.71c	.52b	.40a	.41a	.32	.13	.62c
HavFacAdv*	.58b	.20	.58b	.64c	.45a	.52b	.28	.19	.72c
-IntrUGs	-.70c	-.04	-.71c	-.45b	-.62c	-.67c	-.12	-.17	-.67c
OpnCommcn	.70c	.08	.68c	.56c	.56c	.62c	.30	.21	.80c
StuInput	.72c	.27	.73c	.43b	.46b	.49b	.21	.14	.63c
MaleLess*	.19	.06	.38	.27	.23	.25	.01	.15	.19
MinorMore*	.27	.03	.21	.05	.08	.04	.17	-.09	.27
EnufWmnFc	.31	-.13	.40a	.04	.21	.34a	-.11	.02	.01
EnufEthnF	-.08	.30	.03	.04	-.26	-.06	.39a	.49b	.17
Competitn	-.25	.58b	-.00	.10	-.45b	-.32a	.45b	.49b	.07
AbilMotvn	.62c	.38a	.73c	.34a	.20	.29	.46b	.49b	.71c
A=FcState	.30	-.05	.16	.03	.29	.38a	-.07	.16	.20
TestAccur	.61c	.36	.72c	.48b	.37a	.44b	.29	.17	.50b
GradeVRig	-.27	.33	-.05	-.25	-.58c	-.44b	.24	.24	-.33a
WrkLoad/U	-.13	.68c	.16	-.08	-.65c	-.43b	.55c	.56c	-.08
OfcStaff	.31	-.18	.08	.14	.32	.30	-.01	-.25	.37a
OrientnOK	.49b	-.24	.49b	.43b	.75c	.78c	-.14	.22	.66c
CrseAdvg	.66c	-.11	.58b	.47b	.59c	.63c	.18	.16	.72c
GradInfo	.43b	-.11	.52b	.27	.45b	.49b	.10	.18	.49b
CareerInf	.16	.31	.49a	.36a	.20	.44b	.13	.79c	.40a
TrustFac	.48b	.24	.53b	.35a	.24	.34a	.46b	.41a	.60c
-NtcUGMtg	-.24	.09	-.41a	-.35a	-.57c	-.48b	.06	-.08	-.44b
UGonComm	.48b	-.27	.45a	.42b	.75c	.70c	-.32a	-.03	.36a
TEvalUse	.53c	.38a	.57b	.10	.12	.20	.30	.23	.41a
TEvalGuid	-.02	.04	.09	.11	.30	.37a	-.41b	.21	-.02

See also Table 1 (correlations among outcomes) and Table 3 (likely effects of disciplinary area on aspects of program delivery).

*Unless otherwise indicated, the number of subjects/programs in columns/rows with asterisks is 27 (no biological science programs are included).

a= prob. of significance < .05; b= prob. of significance < .01; c= prob. of significance < .001.

BEST COPY AVAILABLE

Table 3. Aspects of Program Delivery Most Highly Correlated with DiscpArea (Disciplinary Area), the Control Variable (for a technical/scientific versus a human-focus continuum).*

Item Abbv.	Zero-Order Correlation w/DiscpArea	Two-Tailed Probability of Signif.	No.
Wrkload/U	-.82	.000	38
SocIssues	.73	.000	38
-ReqCrses	-.64	.000	38
NonDomToo	.57	.001	29
GradeVRig	-.55	.000	38
UGonComm	.49	.002	38
CrseOpts	.48	.002	38
Competitn	-.46	.004	38
OrientnOK	.42	.008	38
-Explain	-.40	.012	38
EnufEthnF	-.39	.016	38
GradInfo	.37	.023	38
-NtcUGMtg	-.37	.023	38
AdvnCrses	-.33	.044	38
-IntrUGs	-.33	.045	38

*DiscpArea is coded as follows: 1= Engineering and Physical Science Programs; 1.5= Biological/Life-Related Science Programs; 2= Social Science-Related Programs.

Table 4. Aspects of Program Delivery (Levers) Which Served to Promote Program Contribution to Social Responsibility, Community-Useful Skills. This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations				
SocIssues	.82(E)					
OrientnOK	.78	.57	.38	.32	--[a]	--[a]
UGonComm	.70	.47	.40(E)			
-IntrUGs	-.67	--	--	--	--	--
CrseAdvg	.63	.37	.29	.32	--	--
CrseOpts	.63	--	--	--	--	--
OpnCommcn	.62	.30	--	--	--	--
DiscpArea	.59	--	--	--	--	--
StuInput	.49	--	--	--	--	--
GradInfo	.49	--	--	--	--	--
-NtcUGMtg	-.48	-.40	-.28	--	--	--
CareerInf	.44	.60(E)				
GradeVRig	-.44	--	--	--	--	--
TestAccur	.44	.25	(.10)	(.13)	(.24)	.25
-Explain	-.43	(.24)	.28	--	--	--
Wrkload/U	-.43	(-.02)	-.29	--	--	--
EnufSmCls	.43	.33	--	--	--	--
Feedback	.43	--	--	--	--	--
WriteRecn	.41	--	--	--	--	--
OfcStaff	.30	(.11)	.38	.44(E)		
-LDPreReq	(.01)	.26	(.22)	.28	.35(E)	

(Levers) Item Abbv.	Other Zero Order Corr.	No.
NonDomToo	.79	29
FacTeachg	.54	27
HavFacAdv	.52	27
TAEffctvT	.49	26
SfFcCntct	.47	27
EnufFldwk	.47	37

Summaries

	SocIssues	CareerInf	UGonComm	OfcStaff	-LDPrereq
R^2	.667	.786	.820	.855	.872
Adjusted R^2 Total					.852
Sig. F-Change	.0000	.0001	.0163	.0086	.0420

If DiscpArea (Disciplinary Area) is forced to enter first:

	DiscpArea	OrientnOK	SocIssues	CareerInf
R^2	.351	.697	.777	.821
Adjusted R^2 Total				.799
Sig. F-Change	.0001	.0000	.0014	.0077

(E)= Item entered into regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

Table 5. Aspects of Program Delivery (Levers) Which Served to Promote Program Contribution to a Broad, General, Humanistic, or Person-Centered Education. This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations					
SocIssue	.82(E)						
OrientnOK	.75	.49	.29	.35	.30	--[a]	--[a]
UGonComm	.75	.58	.45	.51(E)			
DiscpArea	.74	.34	.28	--	--	--	--
Wrkload/U	-.65	-.46	-.45(E)				
-ReqCrses	-.62	-.37	-.44	-.29	--	--	
-IntrUGs	-.62	(.00)	.29	--	--	--	--
CrseAdvg	.59	.29	(.15)	.31	.41(E)		
GradeVRig	-.58	-.32	-.29	--	--	--	--
-NtcUGMtg	-.57	-.58(E)					
CrseOpts	.57	--	--	--	--	--	--
OpnCommcn	.56	(.17)	(.06)	.29	.37	--	--
-Explain	-.50	--	--	--	--	--	--
StuInput	.46	--	--	--	--	--	--
Competitn	-.45	--	--	--	--	--	--
GradInfo	.45	--	--	--	--	--	--
Feedback	.44	(.04)	(.10)	.25	.39	(.24)	.31
EnufSmCls	.41	.30	(.19)	.28	.29	--	--
WriteRecn	.40	--	--	--	--	--	--
TEvalGuid	.30	.30	(.23)	.31	(.14)	.38(E)	

(Levers) Item Abbv.	Other Zero Order Corr.	No.
NonDomToo	.81	29
FacTeachg	.55	27
HavFacAdv	.45	27
SfFcCntct	.45	27
TAEffctvT	.44	26
EnufFldwk	.40	37

Summaries

	SocIssues	-NtcUGMtg	Wrkload/U	UGonComm	CrseAdvg	TEvalGuid
R^2	.674	.783	.827	.872	.893	.909
Adjusted R^2 Tot.						.891
Sig. F-Change	.0000	.0002	.0056	.0019	.0158	.0299

For DiscpArea (Disciplinary Area) forced to enter first, see Table 6A.

(E)= Item entered into regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

Table 5A. Aspects of Program Delivery (Levers) Which Served to Promote Program Contribution to a Broad, General, Humanistic, or Person-Centered Education, with Disciplinary Area Entered First as a Control.*
This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations					
SocIssue	.82	.60	.35	.28	(.13)	.26	.25	--[a]
OrientnOK	.75	.71(E)						
UGonComm	.75	.65	.49	.41	.42	.40(E)		
DiscpArea	.74(E)							
Wrkload/U	-.65	--	--	--	--	--	--	--
-ReqCrses	-.62	-.28	-.51(E)					
-IntrUGs	-.62	-.59	-.26	--	--	--	--	--
CrseAdvg	.59	.63	(.21)	(.12)	(.00)	(.22)	.34	--
GradeVRig	-.58	-.30	--	--	--	--	--	--
-NtcUGMtg	-.57	-.49	(-.23)	-.28	-.30	-.25	--	--
CrseOpts	.57	.37	--	--	--	--	--	--
OpnCommcn	.56	.69	.33	(.23)	(.07)	.32	.42	.31
-Explain	-.50	-.33	--	--	--	--	--	--
StuInput	.46	.52	--	--	--	--	--	--
Competitn	-.45	--	--	--	--	--	--	--
GradInfo	.45	.28	--	--	--	--	--	--
Feedback	.44	.47	.39	.25	(-.01)	(.22)	.36	.35
EnufSmCls	.41	.63	.39	--	--	--	--	--
WriteRecn	.40	.46	.31	--	--	--	--	--
TestAccu	.37	.52	.48	.42(E)				
OfcStaff	.32	(.23)	(.03)	(.03)	(.15)	.30	.43(E)	
-ProbSolv	-.11	-.25	(.01)	.28	.42(E)			

Summary

	DiscpArea	OrientnOK	-ReqCrses	TestAccu	-ProbSolv	UGonComm	OfcStaff
R^2	.549	.777	.836	.865	.888	.906	.923
Adj. R^2 Total							.905
Sig. F-Change	.0000	.0000	.0014	.0126	.0138	.0217	.0146

*For additional zero-order correlates (with N < 38), see Table 6.

(E)= Item entered into regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

Table 6. Aspects of Program Delivery (Levers) Which Served to Promote Program Contribution to Preparation for a Career.
This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.*	Partial Correlations					
CareerInf	.79(E)						
Wrkload/U	.56	.65	--[a]	--[a]	--[a]	--[a]	--[a]
-ReqCrses	.53	.61	.32	.34(E)			
EnufEthnF	.49	.46	.29	--	--	--	--
AbilMotvn	.49	--	--	--	--	--	--
Competitn	.49	.40	--	--	--	--	--
AdvnCrses	.46	.41	.28	--	--	--	--
DiscpArea	-.45	-.67(E)					
TrustFac	.41	.32	.36(E)----------->{R}				
-Schedule	-.38	-.27	-.30	(-.16)	-.35(E)		

Summaries

	CareerInf	DiscpArea	TrustFac	-ReqCrses	-Schedule	{R}TrustFac
R^2	.630	.797	.824	.846	.865	.858
Adjusted R^2 Tot.					.843	.841
Sig. F-Change	.0000	.0000	.0290	.0386	.0421	.2282

If DiscpArea (Disciplinary Area) is forced to enter first.

	DiscpArea	CareerInf	}	
R^2	.206	.797	}--------->	SAME AS ABOVE
Adj. R^2 Total			}	
Sig. F-Change	.0042	.0000	}	

*There were no additional zero-order correlations ≥ .40.
(E)= Item entered into regression equation at this point.
(R)= Item removed from regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

Table 7. Aspects of Program Delivery (Levers) Which Served to Promote Program Contribution to Preparation for Graduate or Professional School. This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations					
DiscpArea	-.60(E)						
Wrkload/U	.55	--[a]	--[a]	--[a]	--[a]	--[a]	--[a]
AdvnCrses	.55	.46	.38(E)				
TrustFac	.46	.51	.35	.25	.31	--	--
AbilMotvn	.46	.48	.27	--	--	--	--
-ProbSolv	-.45	-.50	-.25	--	--	--	--
Competitn	.45	(.24)	.37	.32	--	--	--
TEvalGuid	-.41	-.39	-.37	-.37	-.31	-.40(E)	
CrseOpts	-.25	(.06)	(-.08)	-.48(E)			
-Explain	(-.18)	-.58(E)					
-NtcUGMtg	(.06)	(-.21)	(-.22)	(-.24)	-.38(E)		

Summary

	DiscpArea	-Explain	AdvnCrses	CrseOpts	-NtcUGMtg	TEvalGuid
R^2	.357	.571	.635	.718	.758	.798
Adj. R^2 Total						.759
Sig. F-Change	.0001	.0002	.0206	.0036	.0287	.0194

(E)= Item entered into regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

(Levers) Item Abbv.	Other Zero Order Corr.	No.
FcDiscsns	.64	27
ChalIntrg	.51	27

Table 8. Aspects of Program Delivery (Levers) Which Served to Promote Program Contribution to Providing Training in Critical Thinking and Analysis. This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations
EnufSmCls	.67(E)	
OpnCommcn	.55	--[a]
WriteRecn	.52	--
TestAccur	.48	--
CrseAdvg	.47	--
-Schedule	-.46	-.29
-IntrUGs	-.45	--
StuInput	.43	--
OrientnOK	.43	--
UGonComm	.42	.31

(Levers) Item Abbv.	Other Zero Order Corr.	No.
NonDomToo	.65	29
HavFacAdv	.64	27
ChalIntrg	.63	27
FcDiscsns	.60	27
FacTeachg	.60	27
SfFcCntct	.50	27
TAEffctvT	.40	26

Summary

	EnufSmCls
R^2	.453
Adj. R^2 Total	.438
Sig. F-Change	.0000

(E)= Item entered into regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

Table 9. Aspects of Program Delivery (Levers) That Correlate with "I will remember a lot of what I have learned in the program's courses for a long time to come." This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations						
StuInput	.72(E)	---------->				(R) .26	--[a]	--[a]
OpnCommcn	.70	.36	.25	--	--	--	--	--
-IntrUGs	-.70	-.27	--	--	--	--	--	--
-Explain	-.69	-.40	-.29	--	--	--	--	--
CrseAdvg	.66	.40	--	--	--	--	--	--
-FrgmtCur	-.65	-.44(E)						
WriteRecn	.65	.39	.35	--	--	--	--	--
Feedback	.65	.31	.26	--	--	--	--	--
SocIssues	.63	.39	.45(E)					
AbilMotvn	.62	.35	(.20)	.31	.37	.43(E)		
TestAccur	.61	.32	.33	.34(E)				
-ProbSolv	-.58	(-.21)	(-.04)	(-.12)	(-.15)	-.25	--	--
-TradMeth	-.55	-.33	--	--	--	--	--	--
TEvalUse	.53	.34	--	--	--	--	--	--
OrientnOK	.49	(.16)	(.11)	(-.15)	(-.10)	(-.03)	-.25	--
UGonComm	.48	.30	.34	(.16)	(.18)	.20	(.13)	.25
TrustFac	.48	--	--	--	--	--	--	--
EnufSmCls	.48	--	--	--	--	--	--	--
-ReqCrses	-.47	-.41	-.34	(-.15)	(-.17)	(-.12)	-.38	-.34
GradInfo	.43	.27	--	--	--	--	--	--
CareerInf	(.16)	(-.08)	(-.10)	(-.09)	(-.17)	(-.13)	-.44(E)	

(Levers) Item Abbv.	Other Zero Order Corr	No
ChalIntrg	.86	27
FacTenchg	.86	27
NonDomTeo	.73	29
SfFcCntct	.63	27
FcDiscsns	.61	27
HavFacAdv	.58	27
TAEffctvT	.54	26

Summary

	StuInput	-FrgmtCur	SocIssues	TestAccu	{R}StuInput	AbilMotvn	CareerInf
R^2	.524	.618	.695	.730	.711	.765	.811
Adj. R^2 Total				.698	.686		.782
Sig. F-Change	.0000	.0059	.0060	.0442	{R}.1334	.0099	.0084

(E)= Item entered into regression equation at this point.

(R)= Item removed from regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

Table 10. Aspects of Program Delivery (Levers) That Correlate with Student Ratings of Overall Satisfaction with Their Major Program "(faculty, teaching, advising, courses/curriculum, testing-grading, governance, etc.)."
This Analysis Is Based on Mean Student Ratings for 38 Academic Major Programs.

(Levers) Item Abbv.	Zero Order Corr.	Partial Correlations								
OpnCommcn	.80(E)									
CrseAdvg	.72	--[a]	--[a]	--[a]	--[a]	--[a]	--[a]	--[a]	--[a]	--[a]
AbilMotvn	.71	.51(E)								
-FrgmtCur	-.70	-.46	-.32	-.36	-.35	-.47(E)				
-IntrUGs	-.67	--	--	--	--	--	--	--	--	--
OrientnOK	.66	--	--	--	--	--	--	--	--	--
EnufSmCls	.63	--	--	--	--	--	--	--	--	--
WriteRecn	.62	--	--	--	--	--	--	--	--	--
TrustFac	.60	--	--	--	--	--	--	--	--	--
-Explain	-.59	--	--	--	--	--	--	--	--	--
SocIssues	.55	(.11)	.25	(-.00)	(.05)	.33	.47(E)			
-Schedule	-.54	(-.17)	(-.10)	(-.17)	(-.13)	(-.02)	.28	.34	--	--
-ProbSolv	-.52	(-.09)	(.12)	(.08)	(.09)	(.07)	(.24)	.26	--	--
Feedback	.51	(-.14)	(-.00)	(.14)	.31	.40	.30	--	--	--
TestAccur	.50	(.20)	(.23)	(.20)	.34	.33	.30	.26	(.13)	.26
GradInfo	.49	(.05)	(-.09)	(-.24)	-.25	(-.17)	-.32	-.42(E)		
-NtcUGMtg	-.44	-.28	--	--	--	--	--	--	--	--
CrseOpts	.42	(.04)	(.08)	(-.03)	(-.03)	(.11)	.30	--	--	--
TEvalUse	.41	--	--	--	--	--	--	--	--	--
CareerInf	.40	.26	--	--	--	--	--	--	--	--
.										
GradeVRig	-.33	-.27	-.54(E)							
.										
A=FcState	(.20)	(-.15)	(-.17)	-.39(E)						
Competitn	(.07)	.30	(.24)	.37	.40(E)					
EnufWmnFc	(.01)	(-.03)	(-.04)	(-.13)	(-.01)	(.09)	(.02)	(-.22)	-.39(E)	

Summary

	OpnCommcn	AbilMotvn	GradeVRig	A=FcState	Competitn	-FrgmtCur	SocIssues	GradInfo	EnufWmnFc
R^2	.644	.737	.814	.842	.867	.896	.919	.934	.944
Adj. R^2 Tot.									.926
Sig. F-Chnge	.0000	.0012	.0006	.0219	.0195	.0060	.0064	.0176	.0342

(E)= Item entered into regression equation at this point.

[a]Figures below .245 are not shown (except, parenthetically, if the partial correlation subsequently rises above that level after having been below it).

(Levers) Item Abbv.	Other Zero Order Corr.	No.
FacTeachg	.86	27
ChalIntrg	.83	27
SfFcCntct	.75	27
HavFacAdv	.72	27
FcDiscsns	.61	27
NonDomToo	.58	29
TAEffctvT	.50	26

References

Astin, A. W. (1980). Developing information systems about student activities and opinions. In L. L. Baird, R. T. Hartnett, & Associates, Understanding student and faculty life (pp. 68-89). San Francisco: Jossey-Bass.

Baird, L. L. (1976). Using self-reports to predict student performance (Research Monograph No. 7, College Entrance Examination Board). Princeton, N.J.: CEEB.

Baird, L. L. (1980a/b). Importance of surveying student and faculty views. And, Using campus surveys for improving colleges. Both in L. L. Baird, R. T. Hartnett, & Associates, Understanding student and faculty life (pp. 1-67 and 198-224, respectively). San Francisco: Jossey-Bass.

Clark, M. J., Hartnett, R. T., and Baird, L. L. (1976). Assessing dimensions of quality in doctoral education: A technical report of a national study in three fields. Princeton, N.J.: Educational Testing Service.

Dressel, P. L. (1980). Improving degree programs: A guide to curriculum development, administration, and review. San Francisco: Jossey-Bass.

Ewell, P. T. (1991). Assessment and public policy: Shifting sands, uncertain future. In Assessment Update, 3, No. 5, pp. 1, 4-7.

Guba, E. G. and Lincoln, Y. S. (1981). Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey-Bass.

Haynes, S. N., and Wilson, C. C. (1979). Behavioral assessment: Recent advances in methods, concepts, and applications. San Francisco: Jossey-Bass.

House, E. R. (1972). The conscience of educational evaluation. Teachers College Record, 73, 405-414.

Institute for Research in Social Behavior (1978). University of California faculty time-use study (Report for 1977-1978 academic year). Oakland, Calif.: Author [456 22nd St., 94612]. (See also reports for 1979 and 1980.)

Jacobson, R. L. (1992, April 15). Colleges face new pressure to increase faculty productivity, in The Chronicle of Higher Education XXXVIII, No. 32, pp. 1, 16-18.

Koon, J. (1977a). An academic review using comparative student evaluations of the life science-related departments and programs at Berkeley. Berkeley, Calif.: Academic Affairs Office, Associated Students of the University of California. (Eric Document Reproduction Service No. ED 165 639)

Koon, J. (1977b). An academic review using comparative student evaluations of selected humanities departments at Berkeley. Berkeley, Calif.: Academic Affairs Office, Associated Students of the University of California. (Eric Document Reproduction Service No. ED 165 638)

Koon, J. (1979). Social science-related programs at Berkeley: An academic review using comparative student evaluations. Berkeley, Calif.: Academic Affairs Office, Associated Students of the University of California. (Eric Document Reproduction Service No. ED 180 903)

Koon, J. (1981a). Engineering and physical science programs at Berkeley: An academic review using comparative student evaluations. Berkeley, Calif.: Academic Affairs Office, Associated Students of the University of California. (Eric Document Reproduction Service No. ED 198 000)

Koon, J. (1981b). The Women's Studies program at Berkeley: An academic review using comparative student evaluations from 15 undergraduate programs in the social sciences area. Berkeley, Calif.: Academic Affairs Office, Associated Students of the University of California. (Eric Document Reproduction Service No. HE 019 796)

Koon, J. (1986, April). The underevaluation of student ratings of academic programs. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (Eric Document Reproduction Service No. ED 271 507)

Koon, J. (1986b, April). Teaching evaluation: Theory and practice for a student position. Issues in Higher Education, 22, 265-274.

Koon, J. (1990). Assessing quality and effectiveness in university and college academic programs: A democratic theory of evaluation (Doctoral dissertation, University of California, Berkeley). Published by University Microfilms Inc. (No. 9126643, 1991), Ann Arbor, Mich.

Koon, J. (1991, March). Some types of outcome-related item wordings in academic program evaluation questionnaires for students. Paper presented at the Second Annual Montclair Assessment Conference ("Confronting Campus Realities and Outcomes Assessment: Making It Work"), Montclair State, N.J.

Larson, J. R. (1979). The limited utility of factor analytic techniques for the study of implicit theories in student ratings of teacher behavior. American Educational Research Journal, 16, No. 2, 201-211.

National Academy of Sciences Committee on Basic Research in the Behavioral and Social Sciences (1982), per the report of Shaw, E. M. (q.v.).

Pace, C. R. (1985). Perspectives and problems in student outcomes research. In P. T. Ewell (Ed.), Assessing educational outcomes (New Directions for Institutional Research, No. 47, pp. 7-18). San Francisco: Jossey-Bass.

Pace, C. R., and Friedlander, J. (1982). The meaning of response categories: How often is "Occasionally," "Often," and "Very Often"? Research in Higher Education, 17, 267-281.

Pascarella, and Terenzini (1991). How College Affects Students. San Francisco: Jossey-Bass.

Shaw, E. M. (1982, July 14). Practical benefits of social-science research cited by National Academy. In The Chronicle of Higher Education, 24, No. 20, p. 8.

Statistical Package for the Social Sciences/Marija J. Norusis (1990). SPSS/PC+ Statistics 4.0. Chicago: SPSS Inc. [444 N. Michigan Ave., IL 60611).

Winkler, A. M. (1992, July 15). Explaining what professors do with their time, "Opinion," The Chronicle of Higher Education, XXXVIII, No. 45, pp. B1-2.

Wise, S. L., Hengstler, D. D., and Braskamp, L. A. (1981). Alumni ratings as an indicator of departmental quality. Journal of Educational Psychology, 73, 71-77.

Appendix A
Item Wordings (1981a only) and
Item Means and Standard Deviations by Disciplinary Area

Outcome Items		SocSci	BioSci	EnggPS	F
Remember: I will remember a lot of what I have learned in the program's courses for a long time to come.	MEAN S.D. No.	3.76 .49 (n=15)	3.66 .29 (n=11)	3.51 .28 (n=12)	1.39
AcadInflu: Please rate the factors below as to their influence on your overall growth and learning while you have been attending college at Berkeley: A. The instructional process: classroom lectures, labs, class assignments (readings, homework, etc.), studying for exams, contact of all kinds with faculty, etc. [1-4 scale, very little to highly.]		3.09 .15		3.32 .16	14.49b
CSubKnow: To what extent are the following goals important to you, and to what extent has your major program met them? F. Gaining a thorough understanding of the subject matter: Program's contribution: [1-5 scale, very low to very high.]		3.83 .34		3.79 .16	.16
CCriticl: D. Providing training in critical thinking and analysis: Program's contribution. [Scale above.]		3.86 .30	3.53 .28	3.86 .20	6.10b
CHumanEd: C. Providing a broad, general, humanistic or person-centered education: Program's contribution:		3.80 .63	2.83 .25	2.63 .18	28.15c
CSocResp: F. Promoting social responsibility or developing skills useful in the community: Program's contribution:		3.39	2.69	2.52	11.13c
CGradSch: B. Preparation for graduate or professional school: Program's contribution:		3.45 .22	3.63 .28	3.85 .20	9.76c
CCarGoal: A. Preparation for a career: Program's contribution:		3.12 .44	3.14 .38	3.60 .32	6.05b
AcadSatis: All in all, how satisfied have you been to date with your major program (faculty, teaching, advising, courses/-curriculum, testing-grading, governance, etc.)? [1-7 scale, all labeled.]		5.21 .71	4.79 .36	5.05 .29	2.11
Curriculum Items (see also Faculty & Instruction)					
-LDPreReqs: Some of the lower division prerequisites are not really necessary for the upper division program.		2.51 .43	2.76 .33	2.81 .53	1.81

		SocSci	BioSci	EnggPS	F
-ReqCrses: There are too many required courses for the major.	MEAN	2.09	2.47	2.78	12.49c
	S.D.				
CrseOpts: There are enough course plan options within the department/program to meet my specialized interests in the subject matter.		3.80 .52	3.57 .34	3.30 .25	5.15a
-FrgmtCur: The courses in this program do not relate well to each other; it is hard to get a good perspective on the field.		2.24 .36	2.22 .32	2.26 .24	0.06
LgLecOK: Large lecture classes can generally provide adequate learning opportunities in this field.		2.45 .38	2.73 .37	2.55 .36	1.80
EnufSmCls: The curriculum is structured so that all majors will have a sufficient number of small classes or seminars.		3.09 .67	2.52 .51	3.14 .50	4.20a
-Schedule: Scheduling of courses in poor in this program: Required courses or subjects of interest to me are offered at the same time, crowded into one or two quarters, or are not being offered this year.		3.07 .40	2.94 .47	3.00 .32	.32
AdvnCrse: Advanced courses are offered in the major (or are otherwise available) in all of my areas of interest.		3.27 .40	3.47 .21	3.51 .27	2.42
NonDomToo: There are enough courses available in this program which explore the contributions of the nondominant "schools of thought" (e.g., environmentalist, low-technology, radical, conservative, futurist, traditionalist, etc.).		3.38 .65		2.58 .30 (n=10)	13.22b
EnufFldWk: There are enough fieldwork, internship, group, lab, or other academic-experiential course offerings in this program, and/or credit is readily available for appropriate types of volunteer experience.		3.22 .62	3.25 .28	3.03 .23 (n=11)	.81
-ProbSolv: The program's methodology, lab, or technical courses do not provide the student with enough experience in real problem-solving and analysis.		2.85 .45	2.68 .33	2.80 .30 (N=11)	.69
-TradMeth: The program's methodology, lab, or technical courses overemphasize traditional methodology and values or are not up-to-date.		2.40 .43	2.09 .25	2.39 .22	3.30a

		SocSci	BioSci	EnggPS	F
-Equipmt: There is a shortage of adequate equipment, computer facilities, supplies, or media materials for use by students in this program.	MEAN S.D.	2.79 (indiv.	2.59 stdt	3.13 averages)	no test
ChalIntrg: The coursework in this program is interesting and challenging.		3.87 .39		3.90 .20	.06
Faculty and Instruction					
Feedback: Comments or review provided by faculty, T.A.s and readers on my midterms, papers, problem sets, experiments and class performance are usually thorough enough.		3.17 .45	3.17 .22	3.03 .29	.72
-Explain: I have needed further explanation or extra instruction in courses in my major, but help or a referral did not seem easy to obtain.		2.48 .33	2.46 .29	2.77 .14	5.03a
SocIssues: Faculty in this program show enough concern about social issues in teaching their subjects.		3.81 .47	3.43 .30	2.92 .17 (n=11)	19.85c
FacTeachg: Teaching by faculty in this program is challenging, of high quality, and increases my interest in the subject matter.		3.73 .46		3.63 .25	.48
TAKnowl: Approximately what percentage of T.A.s (upper or lower division) in your program: Are knowledgeable and well-prepared in the material being taught?		3.79 .22 (n=14)		3.83 .34	.13
TAEffctvT: Approximately what percentage of T.A.s (upper or lower division) in your program: Are effective as teachers in leading class discussions and/or labs?		3.34 .25 (n=14)		3.19 .34	1.72
SfFcCntct: I have been satisfied with the level of interaction I have had with professors here.		3.10 .55		2.97 .27	.54
FcDiscsns: About how many times did you have a discussion outside of the classroom with any professor, apart from study-list advising, in the last three months of instruction (i.e., excluding vacation periods)? [0-4 recoded to actual number of such discussions.]		3.19 .86		3.57 .44	1.92
WriteRec: At least one professor in this program knows me or my work well enough to write me a good recommendation for graduate school or a job.		3.12 .65	2.93 .35	2.98 .41	.52

		SocSci	BioSci	EnggPS	F
-IntrUGs: Faculty members in this program don't really seem interested in or concerned about undergraduates.	MEAN	2.43	2.56	2.76	2.14
	S.D.		.56	.34	.19
OpnCommcn: Communication between faculty and students in this program is open and fairly extensive.		3.45	3.31	3.29	.44
		.66	.35	.34	
StuInput: Most of the faculty in this program are receptive to student input on class structure and content.		3.40	3.35	3.29	.44
		.41	.33	.14	
MaleLess: Generally speaking, faculty in this program seem to be [fill in] supportive of and receptive to male than female students. [1= Much more; [1= Much more ... 5= Much less.]		3.03		3.02	.10
		.18		.13	
MinorMore: Generally speaking, faculty in this program seem to be [fill in] supportive of and receptive to white than ethnic minority students. [1= Much more ... 5= Much less.]		2.91		2.88	.47
		.11		.11	
EnufWmnFc: In my opinion, there are enough women faculty in this program.		2.56	2.84	2.19	3.11
		.58	.91	.28	
EnufEthnF: In my opinion, there are enough ehtnic minority faculty in this program.		2.38	2.45	2.80	3.70a
		.50	.34	.36	
Competition, Grading, Testing					
Competitn: I study and learn best in an environment with competition at a level such as we have here at Berkeley.		2.71	2.67	3.11	7.71b
		.38	.27	.21	
AbilMotvn: I have been impressed by the caliber or motivation of the other students in the major.		3.66	3.63	3.77	.44
		.54	.22	.28	
A=FcState: Faculty in my program clearly state at the outset what the courses will cover and what students need to know or do to get an "A."		3.01	2.98	2.84	1.19
		.25	.40	.18	
TestAccur: Tests given by faculty in most of my courses in this program have measured fairly accurately my knowledge of the material.		3.46	3.38	3.45	.46
		.25	.24	.13	
GradeVRig: Grading in this program is more rigorous than in most--it is very hard to get an "A."		3.30	3.77	3.81	10.05c
		.31	.35	.34	

		SocSci	BioSci	EnggPS	F
Wrkload/U: Compared to work in other programs, the students in this program must do a lot of work for the number of units (credit) awarded.	MEAN	3.00	3.95	4.31	43.20c
	S.D.	.36	.29	.47	
Advising and Information					
OfcStaff: The program's office staff are generally quite helpful.		3.98	3.92	3.77	.92
		.50	.29	.39	
OrientnOK: The program provides a good general orientation for students new to the major.		3.34	2.89	2.93	5.84b
		.53	.28	.18	
CrseAdvg: The program's system of information and advising (faculty, secretaries, catalogs, etc.) makes it easy for me to choose courses, professors, and to understand how best to meet my academic needs.		3.42	3.27	3.17	1.11
		.55	.34	.32	
GradInfo: The faculty and program provide enough information about graduate and professional school programs to which I might want to apply.		3.01	2.98	2.73	3.40a
		.37	.28	.21	
CareerInf: The faculty and program provide enough information about career opportunities for students with a bachelor's degree in this major.		2.76	2.45	2.85	2.25
		.56	.38	.40	
Student Participation					
TrustFac: I trust the faculty to take student needs and views into account in their decision-making for the program.		3.02	3.10	3.09	.25
		.39	.30	.17	
-NtcUGMtg: I have never been notified of a meeting of a general undergraduate organization (or professional society, where equivalent) in this program.		2.00	3.27	2.69	10.47c
		.63	.97	.48	
UGonComm: The program should (even if it already does) include undergraduates on most committees, provide an office, allow a quarterly mailing, etc., to facilitate undergraduate organizing and input.		4.13	3.85	3.87	8.40b
		.24	.20	.14	
TEvalUse: At least one form for student evaluation of teaching was distributed last quarter in each of the program's courses in which I was enrolled.		4.18	4.29	4.19	.62
		.32	.16	.24	
TEvalGuid: All faculty should occasionally use the ASUC Primer questionnaire to evaluate their larger courses and those required for the major, so that students will have access to evaluations of faculty teaching.		4.11	3.97	4.04	1.69
		.17	.20	.20	